Document made available under the Patent Cooperation Treaty (PCT)

International application number: PCT/US05/005254

International filing date: 18 February 2005 (18.02.2005)

Document type: Certified copy of priority document

Document details: Country/Office: US
Number: 60/545,879
Filing date: 19 February 2004 (19.02.2004)

Date of receipt at the International Bureau: 17 March 2005 (17.03.2005)

Remark: Priority document submitted or transmitted to the International Bureau in compliance with Rule 17.1(a) or (b)

World Intellectual Property Organization (WIPO) - Geneva, Switzerland
Organisation Mondiale de la Propriété Intellectuelle (OMPI) - Genève, Suisse

1294351

THE UNITED STATES OF AMERICA

TO ALL TO WHOM THESE PRESENTS SHALL COME:

UNITED STATES DEPARTMENT OF COMMERCE

United States Patent and Trademark Office

March 09, 2005

THIS IS TO CERTIFY THAT ANNEXED HERETO IS A TRUE COPY FROM THE RECORDS OF THE UNITED STATES PATENT AND TRADEMARK OFFICE OF THOSE PAPERS OF THE BELOW IDENTIFIED PATENT APPLICATION THAT MET THE REQUIREMENTS TO BE GRANTED A FILING DATE.

APPLICATION NUMBER: ***60/545,879***
FILING DATE: ***February 19, 2004***
RELATED PCT APPLICATION NUMBER: ***PCT/US05/05254***

UNITED STATES PATENT AND TRADEMARK OFFICE
DEPARTMENT OF COMMERCE

Certified by

Jon W. Dudas

Under Secretary of Commerce
for Intellectual Property
and Director of the United States
Patent and Trademark Office

PROVISIONAL APPLICATION FOR PATENT COVER SHEET

This is a request for filing a PROVISIONAL APPLICATION FOR PATENT under 37 CFR §1.53(c).

Express Mail No. EV334695812US	Docket No. 04-103	Type a plus sign (+) inside this box:	+

INVENTOR(S)/APPLICANTS(S)

LAST NAME	FIRST NAME	MIDDLE INITIAL	RESIDENCE (City and either state or foreign country)
Wang	Avery	L.	Palo Alto, CA

TITLE OF THE INVENTION (280 character maximum)

Method and Apparatus for Identification of Broadcast Source

CUSTOMER NUMBER

20306

McDonnell Boehnen Hulbert & Berghoff

ENCLOSED APPLICATION PARTS (check all that apply)

☒ Specification Number of Pages 8 ☒ Drawings Number of Sheets 2

☒ Other: application Data Sheet

METHOD OF PAYMENT FOR THIS PROVISIONAL APPLICATION FOR PATENT

☐ Applicant claims small entity status. See 37 CFR 1.27 ☒ A check or money order is enclosed to cover the Provisional Filing Fee. ☒ The Commissioner is hereby authorized to charge filing fees and credit Deposit Account Number: **13-2490**.	PROVISIONAL APPLICATION FOR PATENT FILING FEE AMOUNT ($)	160.00

CERTIFICATE OF MAILING

I hereby certify that, under 37 CFR § 1.10, I directed that the correspondence identified above be deposited with the United States Postal Service as "Express Mail Post Office to Addressee," addressed to Mail Stop Provisional Patent Application, Commissioner for Patents, P.O. Box 1450, Alexandria, Virginia 22313-1450, on the date indicated below.

The invention was made by an agency of the United States Government or under a contract with an agency of the United States Government.
_______ No. _______ Yes, the name of the U.S. Government agency and the Government contract number are: _____

Respectfully submitted,

SIGNATURE: [signature] Date: 2/19/04

TYPED or PRINTED NAME Matthew J. Sampson REG. NO. 35,999

☐ Additional inventors are being named on separately numbered sheets attached hereto.

<u>SPECIFICATION</u>
(Case No. 04-103)

<u>TITLE:</u> **Method and Apparatus for Identification of Broadcast Source**

Inventor: Avery Li-Chun Wang
Address: Palo Alto, California

Application Type: Provisional Patent Application

Method and Apparatus for Identification of Broadcast Source

In the field of content identification, sometimes it is desirable to identify not only the content but also the source (such as a channel, stream, or station) of a broadcast transmission. For example, it may be desirable to detect from a free-field audio sample of a radio broadcast which radio station a user is listening to, as well as what song the user is listening to. The technique invented by Wang and Smith, (which is described in International Publication Number WO 02/11123 A2, entitled System and Methods for Recognizing Sound and Music Signals in High Noise and Distortion and claiming priority to US Provisional Application No. 60/222,023 filed July 31, 2000 and US Application Serial No. 09/839,476 filed April 20, 2001, the entire contents of each of which are incorporated herein by reference (hereinafter "Wang and Smith")), can be used to identify free-field audio samples of music playing from various sources, such as radio and television. The Wang and Smith technique may perform a search in a database of audio recordings based on fingerprint hashes extracted from the music. However, because the origin of the audio sample is not relevant to the search, and no broadcast station information is used in the system, it is not easy to determine the exact broadcast station that the use is listening to, if any.

In one embodiment of the system and methods described herein, a user has an audio sampling device containing a **microphone** and optional data transmission means. The user hears an audio program being broadcast from some broadcast means, such as radio or television. He then records a sample using the audio sampling device. The sample is conveyed to an **analyzing means** for analysis to determine which broadcast station the user is listening to. This information may then, for example, be reported back to the user, or combined with an advertisement of a promotion, prize notification, discount offers, and other materials specific for a certain radio station. The information may also be reported to a consumer tracking agency, or otherwise aggregated for statistical purposes. Thus, not only can an audio sample be analyzed to identify its content using a free-field

content identification technique, as described by Wang and Smith, the audio sample may also be analyzed to determine the broadcast source.

Prior art

Watermarks have been used in the past for source identification. Each broadcast station must embed a watermark into the audio stream identifying the station. This technique has several deficiencies. The broadcast station must actively embed the watermark into the audio stream, and furthermore must use a watermarking technique that is an agreed-upon standard used by the source identification system. Any station that does not cooperate by embedding a watermark cannot be identified by these means. Furthermore, the watermark signal must be robust enough to withstand extreme distortion, as is the case if the audio sample is taken in a noisy room with reverberation. Furthermore, in some cases it is desired to use a mobile phone as a sampling device, in which case the audio sample may be subject to lossy compression such as GSM, AMR, EVRC, QCP, etc. In this scenario the audio sample received by the analyzing means may be heavily corrupted and the watermark may not be able to survive such treatment.

Another means that may be used to identify a broadcast station is to perform a cross-correlation analysis between the audio sample and an audio feed captured from the broadcast station (for example from a monitoring station). The matching station should show a strong spike in the cross correlation. A difficulty with cross-correlation is that also in the scenario where a mobile phone is being used to sample the audio and where a lossy compression means is employed, as above. In many voice codecs the phase information is destroyed, and a cross-correlation analysis does not yield a peak even if the audio sample and correct matching broadcast feed are cross-correlated.

Method 1

Use spectrogram peaks, correlate spectrogram peaks rather than direct signal.

Use "combinatorial hash" + peak verification

One technique is to identify a radio station by performing a timestamped recording of a radio broadcast channel, converted into a fingerprint stream. In the field, a user collects a sample, where the sample time is also timestamped in terms of "realtime" offset from a common timebase. A recognition is performed using the technique of Wang and Smith to generate an estimated time offset of the sample within the "original" recording. The absolute times are calculated and compared. If the realtime offsets are within a certain tolerance, say 1 second, then the identification is considered to be originating from the same source, as the probability that a random performance of the same audio content (such as a hit song) is so synchronized in time is extremely low. A rolling buffer of a predetermined length is used to hold a recent fingerprint history. The fingerprints within the rolling buffer are compared against fingerprints generated from the incoming sample. Fingerprints older than a certain cutoff time are ignored, as they are considered to be too far in the past. The length of the buffer is determined by the maximum permissible delay plausible for a realtime simultaneous recording of audio signals originating from a realtime broadcast program, such as network latencies of Voice-over-IP networks, internet streaming, and other buffered content. The delays can range from a few milliseconds to a few minutes.

This may be done by direct comparison of the fingerprint streams (from broadcast channel and from the user) and the relative time offsets. If the relative offset is near zero then it is likely that the streams are being monitored from the same source. Longer and random time delays could mean that the user is listening to an independent but coincident copy of the same audio program.

This method has the attribute of being able to identify the correct broadcast channel without any content identification being required.

One embodiment of a system for implementing method 1 is shown in Figure 1.

Furthermore, fingerprint streams of combinatorial hashes from multiple channels may be grouped into sets of [hash + channel ID + timestamp]. These data structures may be

placed into a rolling buffer ordered by time. The contents of the rolling buffer may further be sorted by hash values for faster search.

The rolling buffer may be instantiated by using batches of time blocks, perhaps M=10 seconds long each: every 10 seconds blocks of new [hash + channel ID + timestamp] are dumped into a big bucket and sorted by hash. Then each block ages, and parallel searches are done for each of N blocks to collect matching hashes, where N*M is the longest history length, and (N-1)*M is the shortest. The hash blocks are retired in conveyor-belt fashion.

The number of matching temporally-aligned hashes is the score.

A further step of verification may be used in which spectrogram peaks may be aligned. Because the Wang/Smith technique generates a relative time offset, it is possible to temporally align the spectrogram peak records within about 10 ms in the time axis. Then we can just count the number of matching time+frequency peaks. That is then the score.

Basically the technique of Wang and Smith but with a rolling buffer of fingerprint values with a time cutoff, dynamically updated.

Method 2: Method via timestamped identification.

A **user audio sample** collected by the user is identified using a **content identification means** such as the one described by Wang and Smith for identifying an audio sample out of a database of audio content files (such as songs). **Broadcast audio samples** are taken periodically taken from each of at least one broadcast channel being monitored by a monitoring station; similarly, a content identification step is performed for each broadcast channel. The broadcast samples must be taken frequently enough so that at least one sample is taken per audio program (i.e. per song) in each broadcast channel. While the user audio sample is collected, a **user sample timestamp** (UST) is taken to mark the beginning time of the audio sample based on a **standard reference clock**.

Similarly for each broadcast sample, a **broadcast sample timestamp** (BST) is also taken to mark the beginning of each sample based on the standard reference clock.

The identification method disclosed by Wang and Smith produces as a consequence of the identification process an accurate relative time offset between the beginning of the identified content file from the database and the beginning of the audio sample being analyzed. Hence, a **user sample relative time offset** (USRTO) and a **user sample identity** are noted as a result of identifying the user audio sample, and a **broadcast sample relative time offset** (BSRTO) and **a broadcast sample identity** is noted as a result of identifying each broadcast audio sample.

The following relations should hold between a user audio sample and a correctly matching broadcast audio sample:

(1) User sample identity = broadcast sample identity AND

(2) UST-USRTO = BST-BSRTO + delay

The delay is a small systematic tolerance that depends on the time difference due to propagation delay of the extra path taken by the user audio sample, for example the latency through a digital mobile phone network.

The probability of misidentification is small, in that a user sample is taken from the wrong broadcast channel or non-monitored audio source (such as a CD player) and happens to satisfy (1) and (2) is fairly small. It is the probability that an independent copy of a song playing on the radio is coincidentally synchronized within a small time delay.

A **decision** is made as to whether the user audio sample originated from a given broadcast source by noting whether (1) and (2) hold. If a broadcast channel is found for which this holds then it is determined that the user is listening to that channel. This information is noted and relayed to a reporting means which uses the information for some follow-on action.

One embodiment of a system for implementing method 2 is shown in Figure 2.

It is noted that the user audio sample may be transmitted to a central identification server, or partially or fully analyzed on the user audio sampling device in order to produce the user sample identity and user sample relative time offset. Furthermore, any algebraic permutation of (2) is within the scope of the invention.

Tracking by common sequencing of broadcast programs

For Methods 1 and 2:

To further verify that the user is actually listening to a given broadcast channel, and that it is not a coincidence, user samples can be taken over a longer period of time, longer than a typical audio program, over a transition between audio programs on the same channel. If it is the correct channel, the content alignment should be continuously maintained. An exception is when the user changes channels. But continuity of identity over a program transition is a strong indicator that the correct broadcast channel is being tracked. Thus we can track equality (1) where (1) continues to hold, but that the sample identity changes, e.g.

(3a) User sample identity[n] = broadcast sample identity[n]

(3b) User sample identity[n+1] = broadcast sample identity[n+1]

(3c) User sample identity[n] ≠ User sample identity[n+1]

where [n] is the nth sample in time.

Tracking score gap patterns within tracks:

When using method 2 to identify music at high duty cycles of sample vs. non-sampled time. Many, if not all broadcast stations incorporate voice over or other non-music material which frequently is superimposed upon the music streams to be identified, ie: DJ's talking over the beginning and end of records. The variations in recognition score, (or indeed non-recognition) constitute a 'signature' of the performance of that track on that station at that time and date, and can be thus used as a further correlation factor to determine station identity.

Method 3: identification enhancement based on derivation of distortion parameters

Another mechanism by which the source identification may be performed is to note certain systematic distortions of the audio as it is being played. As an example, often

times a radio broadcaster will play an audio program slightly faster or slower than the original recording, owing to slight inaccuracies in the crystal oscillator or other timebase used to play back the program recording. The speed percentage stretch of may be measured in the process of identification, for example using the technique of Wang and Culbert (which is described in International Publication No. WO 03/091990 A1, entitled "Robust and Invariant Audio Pattern Matching and claiming priority to US Provisional Application 60/376,055 filed April 25, 2002, the contents of each of which are incorporated herein by reference). If the timebase of the broadcast program is stretched and also substantially similar to the stretch factor measured in the user sample, then the user sample is highly likely to have originated from the same source. To summarize,

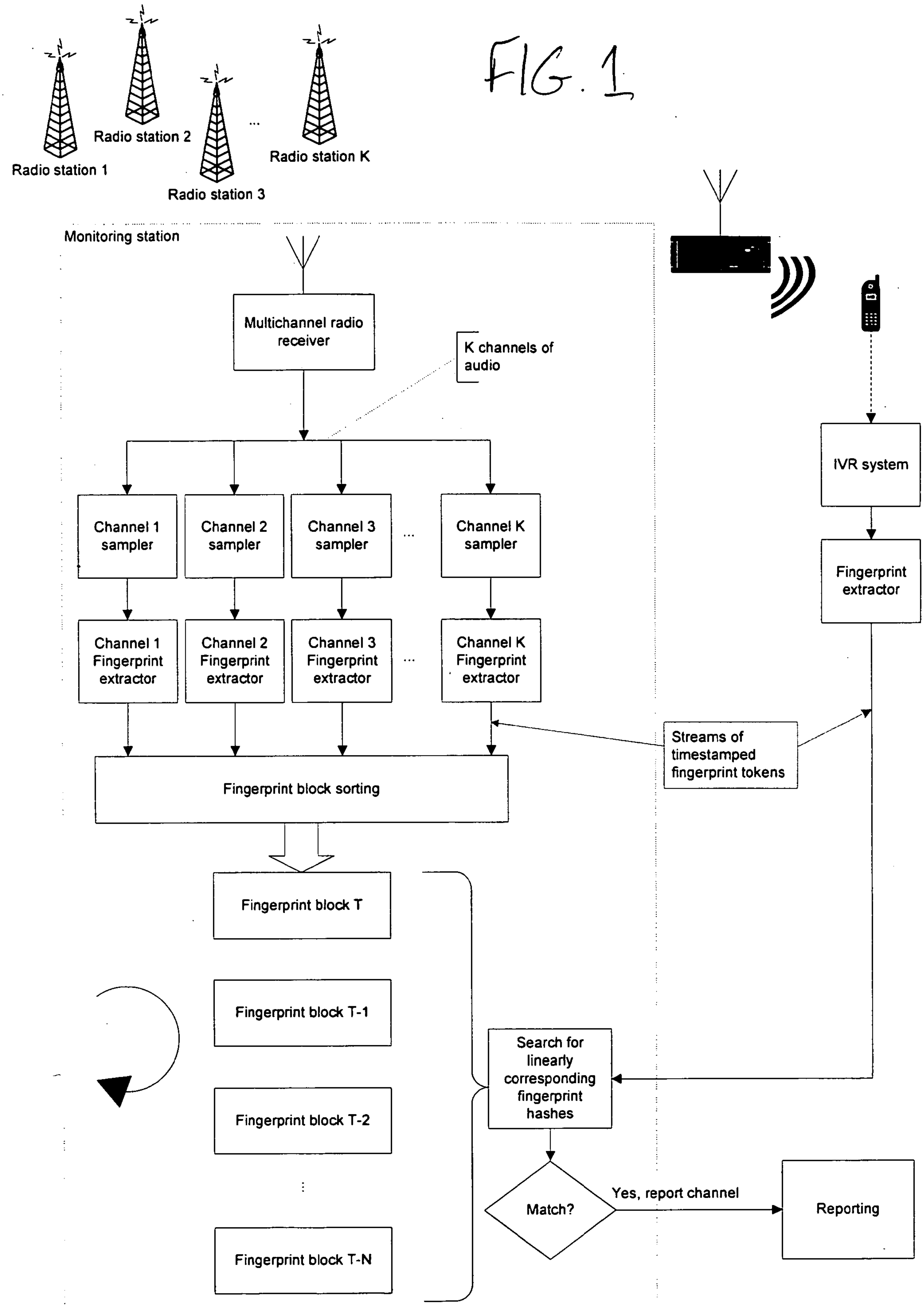
(4) User sample stretch ratio = broadcast sample stretch ratio.

Furthermore, for the purposes of identification, a program may be intentionally stretched by a predetermined amount. The predetermined stretch amount could be used to encode a certain small amount of information. For example, a recording could be stretched to play 1.7% slower. Such a slowdown may not be noticeable to most people. However, if the recognition algorithm is capable of reporting stretch values with 0.05% tolerance, it may be possible to encode 10-20 different messages if playback speeds between –2.0% and +2.0% with 0.1% to 0.2% steps are used.

Furthermore, a stream of information may be embedded in audio by varying the playback speed dynamically (but slowly) over a small range. For example a frame size of 10 seconds could be used, each 10 second segment being sped up or slowed down by a small percentage. If the stretch factors are continually extracted, the values may define a message being sent by the broadcaster.

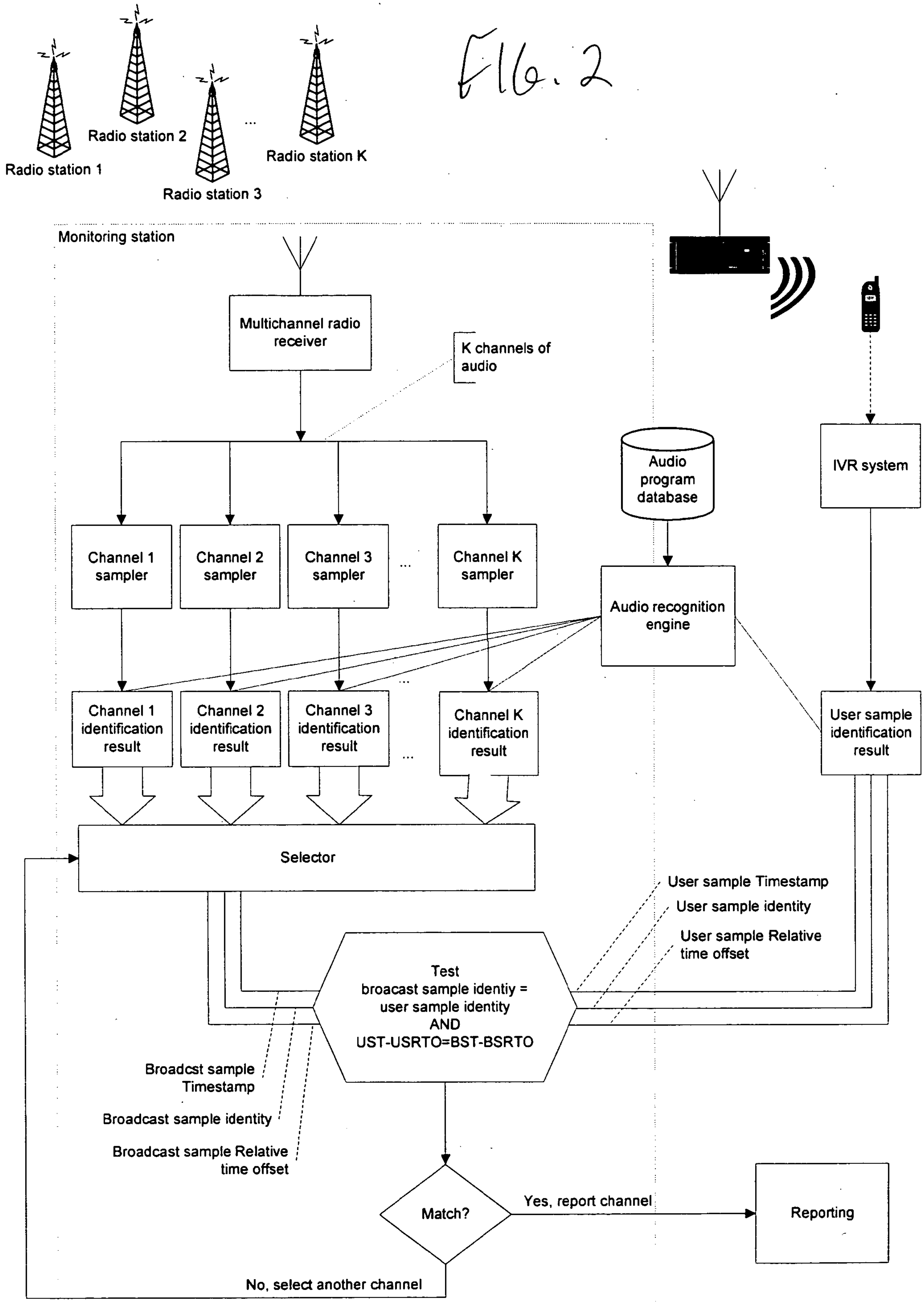
Methods 2 and 3 may be used together to enhance certainty of an opinion that a broadcast channel has been identified.

FIG. 1



Radio station 1
Radio station 2
Radio station 3
...
Radio station K
Monitoring station
Multichannel radio receiver
K channels of audio
Channel 1 sampler
Channel 2 sampler
Channel 3 sampler
...
Channel K sampler
Channel 1 Fingerprint extractor
Channel 2 Fingerprint extractor
Channel 3 Fingerprint extractor
...
Channel K Fingerprint extractor
Streams of timestamped fingerprint tokens
Fingerprint block sorting
Fingerprint block T
Fingerprint block T-1
Fingerprint block T-2
Fingerprint block T-N
IVR system
Fingerprint extractor
Search for linearly corresponding fingerprint hashes
Match?
Yes, report channel
Reporting

FIG. 2



Radio station 1
Radio station 2
Radio station 3
...
Radio station K
Monitoring station
Multichannel radio receiver
K channels of audio
Channel 1 sampler
Channel 2 sampler
Channel 3 sampler
...
Channel K sampler
Audio program database
IVR system
Audio recognition engine
Channel 1 identification result
Channel 2 identification result
Channel 3 identification result
...
Channel K identification result
User sample identification result
Selector
User sample Timestamp
User sample identity
User sample Relative time offset
Test
broacast sample identiy =
user sample identity
AND
UST-USRTO=BST-BSRTO
Broadcst sample Timestamp
Broadcast sample identity
Broadcast sample Relative time offset
Match?
Yes, report channel
Reporting
No, select another channel

Application Data Sheet

Application Information

Application Type::	Provisional
Subject Matter::	
Suggested Classification::	
Suggested Group Art Unit::	
CD-Rom or CD-R?	No
Title::	Method and Apparatus for Identification Of Broadcast Source
Attorney Docket Number::	04-103
Request for Early Publication?::	No
Request for Non-Publication?::	No
Suggested Drawing Figure::	
Total Drawing Sheets::	2
Small Entity::	No
Petition Included?::	No

Applicant Information

Applicant Authority type::	Inventor
Primary Citizenship Country::	US
Status::	
Given Name::	Avery Li-Chun
Family Name::	Wang
City of Residence::	Palo Alto
State of Residence::	CA
Country of Residence::	USA
Street of mailing address::	2925 Ross Road
City of mailing address::	Palo Alto
State or Province of mailing address::	CA
Postal or Zip Code of mailing address::	94303

Correspondence Information

Correspondence Customer Number:: 020306

Representative Information

Representative Customer Number::	020306

Domestic Priority Information

Application::	Continuity Type::	Parent Application::	Parent Filing Date::

Foreign Priority Information

Country::	Application Number::	Filing Date::	Priority Claimed::

Assignee Information

Assignee Name::